社会概念表征和整合的神经基础*

施伟廷 张亚宁 李兴珊 林楠
(中国科学院行为科学重点实验室（中国科学院心理研究所），北京 100101)
(中国科学院大学心理学系, 北京 100049)

摘 要 社会概念的表征和整合是人们理解和思考社会信息的重要基础，其神经基础问题正逐渐成为新兴的研究热点。现有的脑影像和损伤研究证据较为一致地提示，社会概念表征主要依赖由双侧颞叶前部、颞顶联合区、背内侧前额叶、扣带回后部及邻近的楔前叶所组成的脑网络。新近的功能神经影像研究提示，上述脑区除支持社会概念表征外，也分别参与不同层次的社会概念整合。未来研究可以从社会概念表征的子维度、社会概念加工的子成分、社会概念加工的行为影响等方面继续深入和拓展。

关键词 概念表征，概念整合，语言理解，社会认知

收稿日期：2023-03-18
* 国家自然科学基金项目(31871105)资助。
通信作者：林楠, E-mail: linn@psych.ac.cn

人们每天都会通过语言获取大量的社会信息。研究发现，人们的日常交谈约有三分之二是围绕着社会类的话题展开的(Dunbar et al., 1997)。尽管背后涉及十分复杂的认知过程，但通过语言获取社会信息在大多数时候是高效且轻松的——事实上，人们甚至以此为乐，把围绕社会类话题展开的谈话制作成了大量的娱乐节目。人脑是如何高效且轻松地理解语言背后的社会信息的？要探究以上问题的答案，我们便需要了解支撑起社会信息加工的基础——社会概念的表征和整合。

概念构成了词汇含义的心理表征，并允许人们对事物进行分类(Medin & Smith, 1984)。一般认为，大脑在构建复杂意义表征时(例如构建一句话的语义表征)，首先会将概念表征从长时记忆提取到工作记忆中，再对多个概念进行整合。长久以来，人们虽然关注概念表征和整合，却并未将社会概念表征和整合作为一个需要独立研究的问题。直到二十一世纪初，两类重要现象的发现才促使研究者们开展专门针对于社会概念表征和整合的研究。第一类现象是脑区功能的社会领域特殊性现象。研究发现，在大脑的多个认知系统(如视觉、听觉、注意、推理、学习等)中都存在对特定社会类刺激(如面孔、肢体、语音、人格、动机、心理状态)选择性敏感的脑区，这种现象被称作社会领域特殊性现象(Molapour et al., 2021; Spunt & Adolphs, 2017)。一些研究者主张，社会信息加工在物种进化和个体发展的过程中塑造了大脑，形成了所谓“社会脑”系统，从而导致了社会领域特殊性现象(Dunbar & Shultz, 2007; Sallet et al., 2011)。第二类现象是概念在大脑中的分布式表征现象。大量的脑损伤和功能神经影像学证据提示，概念(或称作语义信息)在大脑中的表征是分布式的，不同类型的概念(如形状、颜色、声音、动作、情绪等)的神经表征在脑区分布上有所分离(Binder et al., 2016; Huth et al., 2016; Mahon & Caramazza, 2009)。基于上述两方面的发现，一些研究者推测社会概念的表征和加工可能涉及特殊的神经基础(Caramazza & Mahon, 2006; Mitchell et al., 2002; Zahn et al., 2007; Lin et al., 2015)，从而开启了针对社会概念表征和整合的神经基础研究。

1 社会概念的定义和研究范围

人们对社会概念的定义和研究范围存在一个动态变化的过程。在社会心理学领域，对于人格等特定类型概念的研究由来已久(Allport, 1937)。然而，人们将更加一般性的社会概念作为研究对象并探索其表征和加工是否具有独特的神经基础，却仅有 20 年左右的时间。Mitchell 等(2002)采用功能磁共振(functional magnetic resonance imaging, fMRI)考察人物知识(person knowledge)的神经表征的研究可称作是此类研究的先驱。在该研究中，Mitchell 等人将人物知识定义为“关于社会能动者(social agents)的内隐的、不可观察的属性的一般知识”，

并通过对比人物属性判断任务和物体属性判断任务所诱发的脑激活来揭示人物知识的神经关联。Mitchell 等人所提出的“人物知识”可以被认为是关于社会概念的一类早期的定义。此后，更多的研究者开始从概念表征的角度研究社会知识。不过，尽管这些研究已经将兴趣指向了一般性的社会知识，但从实际采用的实验刺激来看，所考察的范围并未超出社会心理学中的一些经典概念类型，如人格概念(Zahn et al., 2007)和心理状态概念(Mitchell et al., 2005)。

随着社会概念研究的逐渐发展，研究者对社会概念的定义和研究范围有了变化。从定义的角度，新近研究体现出两个特点。第一，新近研究大多从语义维度而非语义类别的角度来定义社会概念(Arioli et al., 2021; Binder et al., 2016; Catricalà et al., 2020; Diveica, et al., 2022; Lin et al., 2015, 2019, 2020; Lin, Wang, et al., 2018; Troche et al., 2017; Vargas & Just, 2020; Villani et al., 2019; Wang et al., 2019; Wang et al., 2023; Zahn et al., 2007; Zhang et al., 2021, 2022)。语义维度的思想起源于多维语义空间假设(Osgood, 1952)，该假设认为概念是沿着多个语义维度进行表征的，每个概念可以被量化表示为多维度语义空间内的一个点。近年来，基于多维语义空间假设所建立的量化语义模型在计算语言学和认知神经科学领域均取得了很大的成功(Binder et al., 2016; Fu et al., 2023; Huth et al., 2016; Kumar, 2021; Wang et al., 2020)。因而，语义维度的思想在当前的概念(语义)研究中正逐渐占据主流。从这一思想出发，新近研究大多假设，在语义空间内存在代表概念社会性(sociality/socialness)高低的连续语义维度，而那些具有较高社会性的概念便是所谓的社会概念，或称作高社会性概念。研究者们倾向于基于语义维度来考察社会概念加工的另一个原因是词汇的意义经常会随着情境发生变化。比如“吃饭”一词虽和社会互动没有必然联系，但在请客、聚餐、婚宴等情境下，“吃饭”便成为了一种社交行为。因此，当综合考虑一个词或概念所适用的全部情境时，人们对不同词汇或概念的社会性评价往往更接近于连续性的分布，而非分类别式的分布。第二，尽管部分早期的社会概念研究强调人物知识的独特性，但新近研究普遍将一个概念在多大程度上涉及人际互动作为衡量它社会性高低的核心标准(Arioli et al., 2021; Binder et al., 2016; Catricalà et al., 2020; Diveica, et al., 2022; Harpaintner et al., 2018; Lin et al., 2015, 2019, 2020; Lin, Wang, et al., 2018; Troche et al., 2017; Vargas & Just, 2020; Wang et al., 2019; Wang et al., 2023; Zhang et al., 2021, 2022)。从实证研究角度，Lin 等(2015)的研究为对比上述两类定义方式的神经现实性提供了重要依据。该研究采用 fMRI 技术，对比了社会动作动词(经常涉及人际互动的人类动作动词，如“扒窃”)、私人动作动词(较少或从不涉及人际互动的人类动作动词，如“刷牙”)、非人类动词(描述物理、化学或生物学现象的动词，如“融化”)的理解所诱发的脑激活。研究结果发现，在双侧颞叶前部、颞顶联合区、背内侧前额叶、扣带回后部及邻近的楔前叶等

脑区，社会动作动词所诱发的激活显著强于其他两类动词；而在其中的大部分脑区，三类动词诱发的激活强度依次显著下降。上述结果提示，大脑对社会概念的敏感程度不仅受到人物知识的影响，更受到概念与人际互动的关联程度的影响。

从研究范围角度，新近研究把研究的范围从社会心理学中的人格、心理状态等经典的概念类型拓展到了物体、动作等更加一般化的概念类型，使得社会概念研究能够直接和经典的概念及语言领域的研究相衔接。在这方面，我国学者 Lin 等(2015, 2019)、Lin, Wang 等(2018)和 Wang 等(2019)所开展的一系列研究颇具代表性。这些研究采用 fMRI 技术，针对过往概念研究中最常见的概念类别，即物体概念、动作概念和抽象概念，考察了不同概念类别下的高社会性词(如“钞票”、“扒窃”、“信赖”)与低社会性词(如“电池”、“刷牙”、“推断”)所诱发神经活动的差别。结果发现，特定脑区的神经活动对于上述不同类别概念的社会性都十分敏感，对高社会性词表现出更强的激活水平。这组发现从神经现实性角度为社会语义维度在不同类型概念表征中的普遍适用性提供了清晰的证据，也使得社会概念研究的实际研究范围得到了较大的拓展。值得注意的是，尽管许多研究仅强调了社会语义维度在抽象概念表征中的重要性(王晓莎, 毕彦超, 2019; Pexman et al., 2022)，但上述研究有力地提示：即便对于最为具体的概念类别——物体概念和动作概念，社会性也是一个有效且重要的语义维度。

综上所述，早期的研究多将社会概念定义为描述内在人物知识的概念类别(如人格、心理状态等)的总和，通过对比社会类别概念与其他类别概念开展研究(如 Mitchell et al., 2002; Zahn et al., 2007)；新近的研究则更多地从语义维度的角度定义社会概念，在同一概念类别内部操控概念的社会性高低(如 Lin et al., 2015, 2019; Wang et al., 2019)，进而逐一考察每个类别概念所包含的社会知识的表征和整合。这种定义方式的改变不仅仅是一种方法学上的变化——在这一改变下，研究者认识到并且证明了，在思考原本被认为是非社会类别(比如无生命的人造物，Lin et al., 2019)的概念时，大脑中仍然存在社会信息加工，并可观测到相应脑区的神经活动。因此，基于语义维度来定义社会概念，能更为全面地反映社会信息在概念系统中的分布。必须强调的是，尽管存在上述定义方式和研究范围的改变，研究者关注社会概念的原因却未曾变化——一直以来，这一主题的研究都源自于对社会概念区别于其他概念的两方面特殊性的关注。第一，社会概念是人们理解和驾驭复杂的社会关系的基石(Pexman et al., 2022)。因此，要从根本上理解人类社会行为背后的认知机制，必然需要对人脑表征和整合社会概念的机理建立深刻的认识。第二，社会概念的表征和加工很可能需要涉及特殊的神经基础，这既符合社会脑假设和概念的分布表征假设的预期(Binder et al., 2016; Binney & Ramsey, 2020)，也正在得到越来越多证据的支持。下文中我们将从社会概念的表征和整合两

个方面对其证据予以综述。

2 社会概念表征的神经基础

社会概念表征的神经基础研究目前主要探究了两个问题：第一，社会概念表征主要涉及哪些脑区和脑网络；第二，社会概念与其他类型概念的脑表征是否相互分离。从研究范式的角度来看，人们多采用词汇理解任务来诱发大脑中社会概念表征的激活，其中，判断词汇之间的语义联系强弱或语义相似性高低是最常用的一类任务(Lin et al., 2015, 2019; Lin, Wang, et al., 2018; Pobric et al., 2016; Wang et al., 2019, 2021; Zahn et al., 2007, 2009)。

关于社会概念表征所涉及的脑区和脑网络问题，存在功能神经影像和损伤两方面的研究证据。功能神经影像研究证据全部来自 fMRI 研究。这其中的大部分研究对比了高社会性概念与低社会性概念在大脑不同区域所诱发的神经激活强度，其背后的基本假设是：参与社会概念表征的脑区应该对高社会性概念表现出更高的激活水平。上文中介绍的 Lin 等(2015, 2019)、Lin, Wang 等(2018)和 Wang 等(2019)的研究便是此类研究的典型例子。目前为止，这一类研究证据已经初具规模，两项针对性的元分析(Arioli et al., 2021; Zhang et al., 2021)都提示，双侧颞叶前部、颞顶联合区、背内侧前额叶、扣带回后部及邻近的楔前叶(见图 1)较为稳定地对高社会性概念表现出比低社会性概念更高的激活水平，提示它们参与了社会概念的加工。以上述研究发现为基础，研究者进一步通过功能连接分析考察了参与社会概念加工的脑区的脑网络属性。研究发现，这些脑区都位于大脑的默认网络(default mode network)之内，并且彼此间具有非常强的功能连接，这提示它们在功能上是紧密联系的，构成了一个支持社会语义加工的脑网络(Lin et al., 2020; Lin, Wang, et al., 2018)。

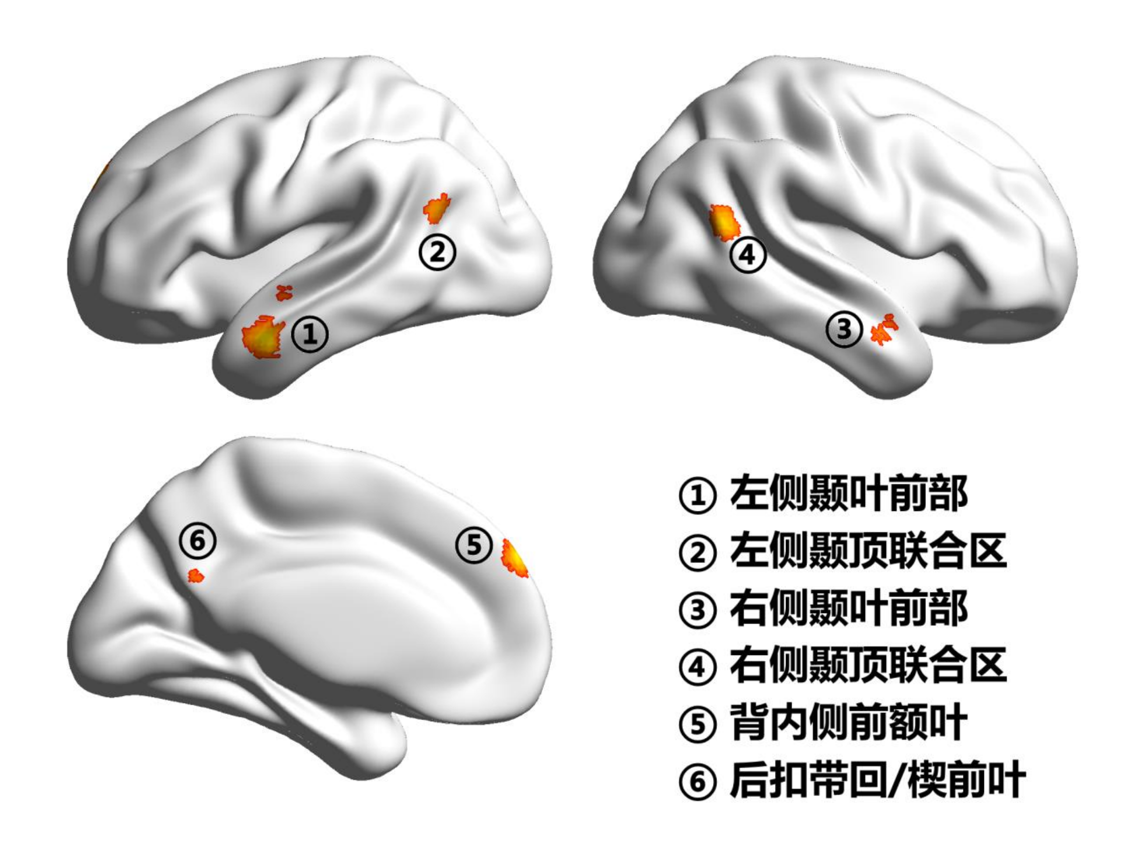


图 1 参与社会概念表征的脑区(资料来源：改编自 Zhang et al., 2021)。

除了考察激活强度的研究，还有少量研究采用多体素模式分析(multivoxel pattern analysis，MVPA)考察社会概念表征的神经关联(Tamir et al., 2016; Thornton & Mitchell, 2018; Van Overwalle et al., 2016)。这类研究所采用的因变量指标是邻近脑区间激活强弱的相对模式。例如，Tamir 等(2016)通过 fMRI 实验获取了 60 个心理状态概念的全脑激活模式，并通过探照灯(searchlight)式的 MVPA 对其影响因素在全脑范围进行了检验。结果发现，心理状态概念在社会性与唤醒度(arousal)的复合维度上的相似性能够预测特定脑区所诱发的多体素激活模式的相似程度，从而提示这部分脑区承担了表征社会情感知识的功能。尽管此类研究数量很少，但其所发现的脑区与前面介绍的基于激活强度的研究基本一致，从而为这些脑区参与社会概念表征提供了另一方面的证据(Tamir et al., 2016; Thornton & Mitchell, 2018; Van Overwalle et al., 2016)。

损伤研究的证据来自脑损伤病人以及对正常被试的经颅磁刺激研究。此类证据的优势在于能够检验脑区功能与行为表现间的因果联系，即特定脑区的功能受到损伤或干扰会否导致社会概念加工困难。但受限于被试的病灶分布或刺激靶点个数，此类研究往往只能检验一部分脑区的功能。已有研究提示了右侧颞叶前部(Catricalà et al., 2020; Pobric et al., 2016; Younes et al., 2022; Zahn et al., 2009, 2017)、左侧顶叶(Wang et al., 2021)、右侧顶内沟(Catricalà et al.,

2020)、左侧颞叶前部、背内侧前额叶、右侧眶额皮层(Zahn et al., 2009)的受损可能诱发社会概念加工困难。对于上述研究的主要发现请见表 1。

表 1 社会概念加工的损伤研究及其主要发现

研究	被试	损伤类型	社会概念加工任务	实验条件	主要发现
Zahn et al., (2009)	额颞叶痴呆患者 29 名；皮质基底综合征患者 18 名；健康对照组被试 30 名	神经退行性疾病	词汇语义关联判断	概念类别：社会或动物	额颞叶痴呆和皮质基底综合征患者均在两类概念上表现出损伤，但额颞叶痴呆患者表现出更严重的社会概念损伤；额颞叶痴呆患者中，右侧颞叶前部损伤患者表现出社会概念选择性损伤；对比社会概念相对动物概念损伤更严重的额颞叶痴呆患者和表现出相反行为模式的额颞叶痴呆患者的脑损伤区域，发现前者在背内侧前额叶、双侧颞叶前部、右侧眶额皮层损伤更重。
Pobric et al., (2016)	健康被试 12 名	虚拟损伤(经颅磁刺激)	词汇语义关联判断	概念类别：社会或动物	对右侧颞叶前部的经颅磁刺激仅显著影响社会概念加工；对左侧颞叶前部的经颅磁刺激同时影响社会和动物概念的加工
	额颞叶痴呆患者 2 名(分别为左、右侧单侧颞叶前部萎缩为主)	神经退行性疾病	词汇语义关联判断	概念类别：社会或动物	左侧和右侧颞叶前部萎缩患者都表现出社会概念损伤和动物概念损伤，但右侧颞叶前部萎缩患者的社会概念相对动物概念损伤程度差异更大。
Zahn et al., (2017)	额颞叶痴呆患者 19 名；健康对照组被试 19 名	神经退行性疾病	任务 1：词汇语义关联判断；任务 2：社会行为后果判断	任务 1：概念类别：社会或动物；任务 2：后果类型：长期后果或短期后果	发现 1 名长期后果知识较社会概念损伤更严重患者，5 名长期后果知识较短期后果知识损伤更严重患者，1 名短期后果较长期后果损伤更严重患者，6 名社会概念较长期后果知识损伤更严重患者；发现额极与右侧颞叶前部损伤程度的相对差异和长期后果知识与社会概念上任务表现的相对差异相关。

Catricalà et al., (2020)	健康被试 36 名	虚拟损伤(经颅磁刺激)	类别启动，即先呈现概念类别标签作为启动词，随后进行概念类别判断	概念类别：社会或数量；启动标签与概念类别的一致性：一致或不一致	对右侧颞叶前部的经颅磁刺激仅影响社会概念类别启动；对右侧顶内沟的经颅磁刺激同时影响社会和数量概念的类别启动
Wang et al., (2021)	脑损伤患者 34 名；健康对照组被试 28 名	多种病因	词汇语义关联判断	概念社会性：社会或非社会；概念情绪性：情绪或非情绪	3 名患者表现出可靠的概念类别分离模式。其中，一名左侧顶叶受损的患者表现出社会概念相对于非社会概念的选择性损伤；一名右侧额叶和颞上皮层受损的患者表现出社会概念相对于非社会概念的选择性幸免；一名左侧额下、脑岛、颞叶前部背侧部受损的患者表现出情绪概念相对于非情绪概念的选择性幸免。

关于社会概念与其他类型概念的脑表征是否相互分离的问题，目前研究存在单分离和双分离两个层次的证据[1]。第一个层次的证据是单分离证据，即可以提示社会概念的表征较之于其他概念的表征涉及额外的神经活动或需要额外的神经结构参与的证据。上文所介绍的研究在实验设计中大都加入了非社会类的概念(基于类别定义社会概念的研究，如 Mitchell et al., 2002; Pobric et al., 2016; Zahn et al., 2007, 2009, 2017)或低社会性概念(基于语义维度定义社会概念的研究，如 Lin et al., 2015, 2019, 2020; Lin, Wang, et al., 2018; Wang et al., 2019, 2021)作为控制条件，并且在数据分析中通过实验条件和控制条件的对比排除了一般性的概念加工效应。对于各个实验所具体采用的实验和控制条件以及主要的研究发现，读者可参见 Zhang 等(2021)文中 Table 2 对脑成像类研究的总结和本文的表 1 对损伤类研究的总结。这些研究的结果都至少构成了单分离证据，能够反映社会概念相对于其他概念所特有的神经活动或神经基础。

第二个层次的证据是双分离证据，即能够提示社会概念的表征与其他概念的表征各自涉及不同的神经活动或需要不同的神经结构参与的证据。双分离证据较之单分离证据其获得条件更为苛刻，实验效应难以被单一的混淆变量(如任务难度)所解释，因此被认为是更为有力的证据。在前文所提及的研究中，有一部分报告了双分离证据。在功能神经影像学研究方面，Mitchell 等(2002)的研究发现，人物知识与物体知识分别在两组不同的脑区诱发更强的神经活动，提示两类知识拥有相互分离的神经表征；Lin, Wang 等(2018)发现大脑中存在不同的区域分别对动词的社会性和视觉表象性敏感，提示社会维度和视觉维度的词汇概念知识在神经表征上有所分离；此外，Wang 等(2019)和 Peer 等(2015)的研究也分别提示了社会概念与情绪概念、空间概念、时间概念在神经表征的脑区分布上彼此分离。在损伤研究方面，Zahn 等(2009)报告了社会概念损伤与动物功能概念损伤的双分离现象，即不同病人分别在社会概念和动物功能概念理解任务上表现出显著更加严重的理解障碍；Wang 等(2021)报道了高社会性抽象概念与低社会性抽象概念损伤的双分离，即不同病人分别在高社会性抽象概念和低社会性抽象概念理解任务上表现出显著更加严重的理解障碍。

以上两个层次的证据较为有力地提示了社会概念表征在神经机制上的特殊性。不过值得注意的是，也有一些研究报告了社会概念与其他概念在神经关联上的重叠。Lin, Wang 等(2018)和 Peer 等(2015)的功能神经影像学研究都发现，尽管社会概念与其他类型概念的神经表征大体上是分离的，但也存在少量重叠区域，主要分布在双侧颞顶联合区和楔前叶，这些重叠的

[1] 本文主要参照 Shallice(1988)改进的单、双分离定义。关于单、双分离定义和逻辑及其演变，请参见 Davies(2010)的论述。

区域被认为很可能是联系和绑定不同类型概念知识的语义中枢(Lin, Wang, et al., 2018)。损伤研究发现，左侧颞叶前部和右侧顶内沟的损伤都会同时导致社会概念与其他概念(如数量概念、动物功能概念)的加工困难(Catricalà et al., 2020; Pobric et al., 2016)。但值得注意的是，功能神经影像数据提示，上述脑区内部的临近区域间存在针对不同类型概念表征的精细功能分离(Lin, Wang, et al., 2018; Wang et al., 2019; Peer et al., 2015)，因此这些脑区损伤所诱发的社会与其他概念同时损伤存在两种可能的原因——它既有可能是因为不同类型概念所依赖的脑区有所重叠，也有可能是因为脑损伤或经颅磁刺激的影响具有一定的空间弥散性，同时影响了彼此相邻的不同功能区域。

此外，另一个值得注意的问题是，在不加以控制的情况下，社会语义维度很容易与其他语义维度发生混淆，这要求我们在解释实验结果时格外小心。幸运的是，这些变量的混淆至少在一部分研究中得到了控制，并且在控制混淆变量后，仍获得了与其他研究相一致的实验发现。比如，Zahn 等(2007)的研究中对比了人格概念和动物功能概念，这一对比混淆了对社会概念的两类定义，即“人物特异性的概念”和“与社会关系和互动相关的概念”。如前文所述，这种定义混淆在 Lin 等(2015)的研究中得到了控制。又如，Tamir 等(2016)的研究首先对多种语义维度进行了主成分分析，从而使得社会性与唤醒度在分析中构成了一个复合维度。因此，其实验结果极有可能混淆了社会概念与情感概念的神经关联。而 Wang 等(2019)和 Arioli 等(2021)专门对社会语义与情感语义的神经关联进行了区分，避免了两者的混淆，且获得了与前人研究相一致的发现。

综上所述，功能神经影像和脑损伤证据较为一致地提示：社会概念的表征依赖一个由双侧颞叶前部、颞顶联合区、背内侧前额叶、扣带回后部及邻近的楔前叶构成的脑网络。该脑网络具有如下特点：第一，网络内部各脑区的激活水平和多体素激活模式均反映出社会概念加工与非社会概念加工的区别；第二，网络内大部分区域的功能对社会概念加工具有特异性，只有少量区域同时也对其他类型的概念敏感；第三，网络内部的脑区之间具有很强的功能连接；第四，网络内局部脑区的损伤将导致社会概念加工困难。

值得注意的是，“社会概念表征依赖一个特异性脑网络”的观点虽曾在 Lin, Wang 等(2018)、Lin 等(2020)、Zhang 等(2021, 2022)的研究中提出，但并未结合多角度的证据加以全面论证。目前为止，在学界占据主流的观点是 Zahn 等(2007)在其经典 fMRI 研究中提出的“社会概念表征在颞叶前部”的论断。对于颞叶前部以外的其他脑区对社会概念表现出的敏感性，多数已有研究或未予讨论，或将推断其可能反映更加高级的社会认知加工(如心理理论)而非社会概念表征本身(Arioli et al., 2021; Binney & Ramsey, 2020; Pexman et al., 2022)，但这些后续研

究自身并未提出新的证据，仅仅是对 Zahn 等(2007)观点的沿袭。然而，Zahn 等(2007)的发现并不能否定颞叶前部以外的其他脑区的社会概念表征功能。Zahn 等(2007)提出其论断所依据的证据来自对 3 类激活强度效应的联合分析，分别为社会类别效应(社会概念强于动物功能概念)、行为描述性效应(概念越能够描述特定社会行为则诱发激活越强)和语义相关效应(语义相关词对诱发的激活强于无关词对)。在这项联合分析中，仅右侧颞叶前部在全脑分析中显示出显著结果。该结果虽然加强了右侧颞叶前部参与社会概念表征的证据，但并不能用来否定其他脑区的社会概念表征功能。从理论角度，我们没有理由假定所有参与社会概念表征的脑区必须表现出行为描述性效应和语义相关效应；从方法学角度，我们不应基于统计上不显著的结果做出推断。通过前面对于功能神经影像和脑损伤证据的全面总结，本文发现颞叶前部之外的其他多个脑区(颞顶联合区和大脑中线结构)对社会概念表征的重要性得到了多方面(激活强度、多体素激活模式、损伤对行为的影响)的证据支持。尤其值得注意的是，采用 MVPA 的研究普遍发现了这些脑区表现出了社会概念解码效应(Tamir et al., 2016; Thornton & Mitchell, 2018; Van Overwalle et al., 2016)——基于这些脑区的激活，研究者甚至可以有效解码出句子语义中所蕴含的特定人格维度信息(Thornton & Mitchell, 2018; Van Overwalle et al., 2016)。这类证据较之 Zahn 等(2007)所考察的行为描述性效应和语义相关效应更为直接地反映了社会概念表征的神经基础。因此，我们认为目前的研究证据更为支持“社会概念表征依赖一个特异性脑网络”的观点。

3 社会概念整合的神经基础

概念的整合包含多个层次。对于语言理解来说，将词汇整合为短语，将短语整合为句子，将句子整合为语篇，需要不同层次的概念整合。对概念整合的一般性神经基础研究由来已久，这些研究发现左侧额下回等脑区是参与一般性的概念(语义)整合的关键脑区(朱祖德 等, 2011)。然而，在概念表征的研究中，人们发现不同类型的概念信息(如视觉、运动、社会、情感、时间、空间等)分布式地表征在不同的脑区和神经网络中(Binder et al., 2016; Wang et al., 2023)，大脑中这种分布式的概念表征方式在传统的概念整合研究中并未引起足够的关注。概念的表征与整合是彼此紧密联系的认知功能，并且基于一些新近流行的理论模型(如加工记忆模型，Hasson et al., 2015)，信息的表征和加工往往依赖相同的脑区。据此推论，概念整合也有可能发生在分布式的概念表征系统中，或者与之紧密联系的脑区中，即不同类型的概念信息的整合很可能依赖于不同脑区和神经网络。目前为止，考察特定类型概念整合的神经关联的研究还很少。此类研究沿用考察一般性的概念(语义)整合的基本研究范式，并通过操

控刺激材料所包含的概念信息类型，对比不同类型概念整合在神经关联上的异同。初步的研究证据提示，社会概念整合相较于非社会概念整合有独特的神经关联，且在脑区分布上与社会概念表征研究所发现的脑区十分相似。

两项 fMRI 研究考察了短语水平的社会概念整合的神经基础。Lin 等(2020)通过操纵短语的社会性和合理性来考察短语水平的社会概念整合，其观测指标为脑区的激活水平。研究结果发现，存在不同脑区分别对短语的社会性和合理性敏感，提示这些脑区分别参与了社会语义加工和短语语义整合。然而，Lin 等人并未发现两个因素之间在任何脑区上存在显著的交互作用，因此未能提示任何脑区选择性参与短语的社会概念整合。Yang 和 Bi(2022)使用了 5 类刺激来考察短语水平的社会概念整合，分别为社会短语(如拥抱母亲)、操作短语(如折叠纸巾)、不匹配短语(如折叠病患)、社会词(如拥抱)、操作词(如折叠)。该研究通过三类指标反映社会概念整合，包括社会短语的激活强度结果与其他条件的差异、社会短语的表征相似性分析结果与其他条件(操作短语、社会词、操作词)的差异、社会短语与对应词汇激活的加法和乘法相关与操作短语与对应词汇激活的加法和乘法相关的差异。研究结果发现，在枕下回和临近的颞下回区域，上述三类指标都提示了社会短语整合效应。Yang 和 Bi 提出，该区域的发现虽一定程度上出乎预料，但该区域临近枕叶面孔区，且曾被发现参与一些高级社会认知任务，因此可能至少参与了部分类型的社会概念的整合。另外，双侧颞叶前部也在表征相似性分析中表现出了社会概念整合效应。

另有两项研究考察了句子和语篇水平的社会概念整合的神经基础。Zhang 等(2021)通过一项 fMRI 实验同时考察了句子和语篇水平的社会概念整合的神经基础。该实验同时操纵了刺激的语言层级(语篇、句子、词)和语义社会性(高、低)，并通过两类独立的指标来反映社会概念整合，即不同语言层级上语义社会性效应的差异，以及在分层回归中排除刺激在较低层级语言单位上的语义社会性评分的贡献后，刺激在高层级语言单位上的社会性评分对刺激所诱发激活的解释力。两类指标均提示，参与社会概念加工的经典脑区——双侧颞叶前部、颞顶联合区、背内侧前额叶、扣带回后部及邻近的楔前叶都参与了句子层级的社会概念整合。此外，分层回归结果还提示，双侧颞顶联合区可能参与了语篇水平的社会概念整合。Lin, Yang 等(2018)的研究通过对比社会类和非社会类语篇中开头句子与结尾句子诱发的脑激活差异考察了语篇水平的社会概念整合。该对比背后的假设是结尾句子需要整合前文语境，因此涉及更多的语义整合加工。研究结果发现，双侧颞顶联合区以及颞中回都在社会类语篇中表现出较之非社会类语篇更强的社会概念整合效应，并且右侧的颞顶联合区和颞中回所表现出的语义整合效应仅特异性地存在于社会类语篇加工中。此外，Kaplan 等(2017)也报告了与 Lin

等人相似的发现，即在双侧颞顶联合区以及额叶和顶叶内侧区域，语篇结尾部分的社会语义效应较之开头部分都更强。

一个重要的问题是，社会概念整合的研究发现是否可以类比推广到其他类型的概念整合的研究中？目前为止，尚缺乏足够的研究证据可以为我们清晰提示非社会类概念(如视觉、运动)的整合的神经基础。一方面，在文献中对此类问题的直接考察非常少，且均报告了阴性结果(Price et al., 2015; Yang & Bi, 2022)。另一方面，有的语篇加工研究报告了不同类型语篇信息在大脑中的分布式表征和加工的证据(Speer et al., 2009; Tamir et al., 2016)，但这些研究并未控制词汇水平的概念激活效应，所以并不能证明非社会类的概念整合的神经关联是分布式的。因此，非社会类概念整合的神经基础问题仍有待进一步探索。

综上所述，对比社会概念表征与社会概念整合研究的结果，我们发现：社会概念表征研究所发现的主要脑区都参与了一个或多个层次的社会概念整合，其中短语水平的社会概念整合可能涉及颞叶前部，语篇水平的社会概念整合主要涉及颞顶联合区，而句子水平的社会概念整合则可能涉及了全部脑区。在上述脑区之外，双侧枕下回和颞下回可能参与了短语水平的社会概念整合，双侧颞中回可能参与了语篇水平的社会概念整合。但需要注意的是，目前针对社会概念整合的研究数量较少，均依赖于单一的技术(fMRI)，且仅发现了单分离(即社会语义整合所额外诱发的神经活动信号)的证据，因此这类研究的结论还有待进一步的检验。

4 展望

社会概念表征和整合的研究方兴未艾，它在很多方面都值得进一步探索。其中，笔者认为，在接下来的研究阶段，以下五个问题显得尤为突出。

第一，社会概念表征的子维度。社会概念所涵盖的信息类型十分丰富。从语义维度的角度出发，目前许多社会概念表征的神经基础研究只对社会概念与非社会概念进行粗略区分，但必然还存在更加细致的社会语义子维度可以对不同内容的社会概念进行更加细致的区分。这一点在人格等特定社会概念领域已经被大量研究证实(Thornton & Mitchell, 2018)。未来，对社会概念维度的全面和深入探索必将展开。

第二，社会概念加工的子成分。概念加工不仅涉及整合，还涉及到认知控制、工作记忆、推理等多方面的认知成分。从社会脑假设出发，在以上环节都有可能存在社会领域特殊性现象。针对于其中一些认知成分，已经有研究者提出了此类假设(如社会控制，对此问题的综述请参见 Binney & Ramsey, 2020)。除了社会概念的表征和整合，社会概念加工还包含哪些特殊的子成分？这也是未来社会概念加工研究所必须回答的问题。

第三，社会概念加工所依赖的脑网络内的精细分工。通过近年来的研究，我们已经知道，社会概念表征和整合依赖多个脑区，且这些脑区间存在紧密的功能连接，形成了一个脑网络。然而，我们对于其内部的脑区间精细分工尚缺乏足够的了解。社会语义整合的研究初步揭示了上述脑区间的功能差异，但此类研究证据还非常稀缺，对结果解释的推论性较高，有待进一步深入检验和探索。

第四，澄清过往功能神经影像研究所揭示的语言和语义加工的一般性神经基础是否受到了社会语义加工效应的混淆。自然言语中很大比例围绕社会类话题展开(Dunbar et al., 1997)，因此自然的语言刺激大多包含较为丰富的社会概念信息，而这些社会概念所诱发的神经激活很可能被混淆为一般性的语言或语义加工所诱发的神经激活。与这一推测相一致，过往文献所揭示的语言加工脑网络(Malik-Moraleda et al., 2022)和一般性语义表征脑网络(Binder et al., 2009)都包含大量对社会概念敏感的脑区。基于目前已有的文献(如 Huth et al., 2016; Lin et al., 2020; Lin, Wang, et al., 2018)，人们已经知道，许多所谓一般性语言或语义脑网络中的脑区并非对所有类型的概念信息都同等程度敏感，这些脑网络包含了对各种类型概念(如社会概念、视觉概念)相对敏感的脑区。那么，过往功能神经影像研究所揭示的语言加工脑网络和一般性语义表征脑网络是否受到了社会语义加工效应的混淆？这些脑网络中的一部分脑区是否仅对社会概念的加工敏感？这是一个亟待澄清的问题。

第五，社会概念加工的行为影响。不同类型概念在表征和加工机制上的差异有可能导致其在行为任务中的差异，比如高表象性的词和高情绪性的词在词汇判断等任务上会表现出相对较短的反应时(Balota et al., 2004; Kousta et al., 2011)。社会概念的表征和整合具有相对独特的神经机制，这在行为层面上会有怎样的影响？目前这方面的研究目前刚刚起步(Diveica et al., 2022; Xia et al., 2023)。未来，行为与脑两方面研究证据的结合将更加有利于我们揭示社会概念表征和整合的认知神经机制。

参考文献

王晓莎，毕彦超. (2019). 抽象概念语义表征的认知神经基础研究. *生理学报, 71*(1), 117–126.

朱祖德，王穗苹，冯刚毅，刘颖. (2011). 左侧额下回在句子语义整合加工中的作用. *心理科学进展, 19*(8), 1147–1157.

Allport, G. W. (1937). Personality: A psychological interpretation. *American Journal of Sociology, 45*(1), 48–50.

Arioli, M., Basso, G., Poggi, P., & Canessa, N. (2021). Fronto-temporal brain activity and connectivity track implicit attention to positive and negative social words in a novel socio-emotional Stroop task. *Neuroimage, 226*(4), 117580.

Arioli, M., Gianelli, C., & Canessa, N. (2021). Neural representation of social concepts: A coordinate-based meta-analysis of fMRI studies. *Brain Imaging and Behavior, 15*, 1912–1921.

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General, 133*(2), 283–316.

Binder, J. R., Conant, L. L., Humphries, C. J., Fernandino, L., Simons, S. B., Aguilar, M., & Desai, R. H. (2016). Toward a brain-based componential semantic representation. *Cognitive Neuropsychology, 33*(3-4), 130–174.

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex, 19*(12), 2767–2796.

Binney, R. J., & Ramsey, R. (2020). Social semantics: The role of conceptual knowledge and cognitive control in a neurobiological model of the social brain. *Neuroscience & Biobehavioral Reviews, 112*, 28–38.

Caramazza, A., & Mahon, B. Z. (2006). The organisation of conceptual knowledge in the brain: The future's past and some future directions. *Cognitive Neuropsychology, 23*(1), 13–38.

Catricalà, E., Conca, F., Fertonani, A., Miniussi, C., & Cappa, S. F. (2020). State-dependent TMS reveals the differential contribution of ATL and IPS to the representation of abstract concepts related to social and quantity knowledge. *Cortex, 123*, 30–41.

Davies, M. (2010). Double dissociation: Understanding its role in cognitive neuropsychology. *Mind & Language, 25*(5), 500–540.

Diveica, V., Pexman, P. M., & Binney, R. J. (2022). Quantifying social semantics: An inclusive definition of socialness and ratings for 8388 English words. *Behavior Research Methods, 55*(2), 461–473.

Dunbar, R. I., Marriott, A., & Duncan, N. D. (1997). Human conversational behavior. *Human Nature, 8*(3), 231–246.

Dunbar, R. I., & Shultz, S. (2007). Evolution in the social brain. *Science, 317*(5843), 1344–1347.

Fu, Z., Wang, X., Wang, X., Yang, H., Wang, J., Wei, T., ... Bi, Y. (2023). Different computational relations in language are captured by distinct brain systems. *Cerebral Cortex, 33*(4), 997–1013.

Harpaintner, M., Trumpp, N. M., & Kiefer, M. (2018). The semantic content of abstract concepts: A property listing study of 296 abstract words. *Frontiers in Psychology, 9*, 1748.

Hasson, U., Chen, J., & Honey, C. J. (2015). Hierarchical process memory: Memory as an integral component of information processing. *Trends in Cognitive Sciences, 19*(6), 304–313.

Huth, A. G., De Heer, W. A., Griffiths, T. L., Theunissen, F. E., & Gallant, J. L. (2016). Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature, 532*(7600), 453–458.

Kaplan, J. T., Gimbel, S. I., Dehghani, M., Immordino-Yang, M. H., Sagae, K., Wong, J. D., ... Damasio, A. (2017). Processing narratives concerning protected values: A cross-cultural investigation of neural correlates. *Cerebral Cortex, 27*(2), 1428–1438.

Kousta, S. T., Vigliocco, G., Vinson, D. P., Andrews, M., & Del Campo, E. (2011). The representation of abstract words: Why emotion matters. *Journal of Experimental Psychology: General, 140*(1), 14–34.

Kumar, A. A. (2021). Semantic memory: A review of methods, models, and current challenges. *Psychonomic Bulletin & Review, 28*, 40–80.

Lin, N., Bi, Y. C., Zhao, Y., Luo, C. M., & Li, X. S. (2015). The theory-of-mind network in support of action verb comprehension: Evidence from an fMRI study. *Brain and Language, 141*, 1–10.

Lin, N., Wang, X., Xu, Y., Wang, X., Hua, H., Zhao, Y., & Li, X. (2018). Fine subdivisions of the semantic network supporting social and sensory-motor semantic processing. *Cerebral Cortex, 28*(8), 2699–2710.

Lin, N., Xu, Y., Wang, X., Yang, H., Du, M., Hua, H., & Li, X. (2019). Coin, telephone, and handcuffs: Neural correlates of social knowledge of inanimate objects. *Neuropsychologia, 133*, 107187.

Lin, N., Xu, Y., Yang, H., Zhang, G., Zhang, M., Wang, S., ... Li, X. (2020). Dissociating the neural correlates of the sociality and plausibility effects in simple conceptual combination. *Brain Structure and Function, 225*, 995–1008.

Lin, N., Yang, X., Li, J., Wang, S., Hua, H., Ma, Y., & Li, X. (2018). Neural correlates of three cognitive processes involved in theory of mind and discourse comprehension. *Cognitive, Affective, & Behavioral Neuroscience, 18*, 273–283.

Mahon, B. Z., & Caramazza, A. (2009). Concepts and categories: A cognitive neuropsychological perspective. *Annual Review of Psychology, 60*(1), 27–51.

Malik-Moraleda, S., Ayyash, D., Gallée, J., Affourtit, J., Hoffmann, M., Mineroff, Z., ... Fedorenko, E. (2022). An

investigation across 45 languages and 12 language families reveals a universal language network. *Nature Neuroscience, 25*(8), 1014–1019.

Medin, D. L., & Smith, E. E. (1984). Concepts and concept formation. *Annual Review of Psychology, 35*(1), 113–138.

Molapour, T., Hagan, C. C., Silston, B., Wu, H., Ramstead, M., Friston, K., & Mobbs, D. (2021). Seven computations of the social brain. *Social Cognitive and Affective Neuroscience, 16*(8), 745–760.

Mitchell, J. P., Heatherton, T. F., & Macrae, C. N. (2002). Distinct neural systems subserve person and object knowledge. *Proceedings of the National Academy of Sciences, 99*(23), 15238–15243.

Mitchell, J. P., Neil Macrae, C., & Banaji, M. R. (2005). Forming impressions of people versus inanimate objects: Social-cognitive processing in the medial prefrontal cortex. *Neuroimage, 26*(1), 251–257.

Osgood, C. E. (1952). The nature and measurement of meaning. *Psychological Bulletin, 49*(3), 197–237.

Peer, M., Salomon, R., Goldberg, I., Blanke, O., & Arzy, S. (2015). Brain system for mental orientation in space, time, and person. *Proceedings of the National Academy of Sciences, 112*(35), 11072–11077.

Pexman, P. M., Diveica, V., & Binney, R. J. (2022). Social semantics: The organization and grounding of abstract concepts. *Philosophical Transactions of the Royal Society B, 378*(1870), 20210363.

Pobric, G., Lambon Ralph, M. A., & Zahn, R. (2016). Hemispheric specialization within the superior anterior temporal cortex for social and nonsocial concepts. *Journal of Cognitive Neuroscience, 28*(3), 351–360.

Price, A. R., Bonner, M. F., Peelle, J. E., & Grossman, M. (2015). Converging evidence for the neuroanatomic basis of combinatorial semantics in the angular gyrus. *Journal of Neuroscience, 35*(7), 3276–3284.

Sallet, J., Mars, R. B., Noonan, M. P., Andersson, J. L., O'reilly, J. X., Jbabdi, S., ... Rushworth, M. F. (2011). Social network size affects neural circuits in macaques. *Science, 334*(6056), 697–700.

Shallice, T. (1988). *From neuropsychology to mental structure*. Cambridge University Press.

Speer, N. K., Reynolds, J. R., Swallow, K. M., & Zacks, J. M. (2009). Reading stories activates neural representations of visual and motor experiences. *Psychological Science, 20*(8), 989–999.

Spunt, R., & Adolphs, R. (2017). A new look at domain specificity: Insights from social neuroscience. *Nature Reviews Neuroscience, 18*(9), 559–567.

Tamir, D. I., Bricker, A. B., Dodell-Feder, D., & Mitchell, J. P. (2016). Reading fiction and reading minds: The role of simulation in the default network. *Social Cognitive and Affective Neuroscience, 11*(2), 215–224.

Tamir, D. I., Thornton, M. A., Contreras, J. M., & Mitchell, J. P. (2016). Neural evidence that three dimensions organize mental state representation: Rationality, social impact, and valence. *Proceedings of the National*

Academy of Sciences, 113(1), 194–199.

Thornton, M. A., & Mitchell, J. P. (2018). Theories of person perception predict patterns of neural activity during mentalizing. *Cerebral Cortex, 28*(10), 3505–3520.

Troche, J., Crutch, S., & Reilly, J. (2017). Defining a conceptual topography of word concreteness: Clustering properties of emotion, sensation, and magnitude among 750 English words. *Frontiers in Psychology, 8*, 1787.

Van Overwalle, F., Ma, N., & Baetens, K. (2016). Nice or nerdy? The neural representation of social and competence traits. *Social Neuroscience, 11*(6), 567–578.

Vargas, R., & Just, M. A. (2020). Neural representations of abstract concepts: Identifying underlying neurosemantic dimensions. *Cerebral Cortex, 30*(4), 2157–2166.

Villani, C., Lugli, L., Liuzza, M. T., & Borghi, A. M. (2019). Varieties of abstract concepts and their multiple dimensions. *Language and Cognition, 11*(3), 403–430.

Wang, X., Li, G., Zhao, G., Li, Y., Wang, B., Lin, C. P., ... Bi, Y. (2021). Social and emotion dimensional organizations in the abstract semantic space: The neuropsychological evidence. *Scientific Reports, 11*(1), 23572.

Wang, X., Wang, B., & Bi, Y. (2019). Close yet independent: Dissociation of social from valence and abstract semantic dimensions in the left anterior temporal lobe. *Human Brain Mapping, 40*(16), 4759–4776.

Wang, S., Zhang, Y., Shi, W., Zhang, G., Zhang, J., Lin, N., & Zong, C. (2023). A large dataset of semantic ratings and its computational extension. *Scientific Data, 10*(1), 106.

Wang, S., Zhang, J., Wang, H., Lin, N., & Zong, C. (2020). Fine-grained neural decoding with distributed word representations. *Information Sciences, 507*, 256–272.

Xia, A., Solomon, S. H., Thompson-Schill, S. L., & Jenkins, A. C. (2023). Constructing complex social categories under uncertainty. *Cognition, 234*, 105363.

Yang, H., & Bi, Y. (2022). From words to phrases: Neural basis of social event semantic composition. *Brain Structure and Function, 227*(5), 1683–1695.

Younes, K., Borghesani, V., Montembeault, M., Spina, S., Mandelli, M. L., Welch, A. E., ... Gorno-Tempini, M. L. (2022). Right temporal degeneration and socioemotional semantics: Semantic behavioural variant frontotemporal dementia. *Brain, 145*(11), 4080–4096.

Zahn, R., Green, S., Beaumont, H., Burns, A., Moll, J., Caine, D., ... Ralph, M. A. L. (2017). Frontotemporal lobar degeneration and social behaviour: Dissociation between the knowledge of its consequences and its conceptual meaning. *Cortex, 93*, 107–118.

Zahn, R., Moll, J., Iyengar, V., Huey, E. D., Tierney, M., Krueger, F., & Grafman, J. (2009). Social conceptual

impairments in frontotemporal lobar degeneration with right anterior temporal hypometabolism. *Brain, 132*(3), 604–616.

Zahn, R., Moll, J., Krueger, F., Huey, E. D., Garrido, G., & Grafman, J. (2007). Social concepts are represented in the superior anterior temporal cortex. *Proceedings of the National Academy of Sciences, 104*(15), 6430–6435.

Zhang, G., Hung, J., & Lin, N. (2022). Coexistence of the social semantic effect and non-semantic effect in the default mode network. *Brain Structure and Function, 228*(4), 321–339.

Zhang, G., Xu, Y., Zhang, M., Wang, S., & Lin, N. (2021). The brain network in support of social semantic accumulation. *Social Cognitive and Affective Neuroscience, 16*(4), 393–405.

Neural basis of social concept representation and social semantic integration

SHI Weiting, ZHANG Yaning, LI Xingshan, LIN Nan

(*CAS Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Sciences, Beijing* 100101, *China*)

(*Department of Psychology, University of Chinese Academy of Sciences, Beijing* 100049, *China*)

Abstract: The representation and integration of social concepts is the basis of social semantic comprehension and social thinking. In recent years, it becomes a new research hotspot, bridging the gap between the neuroscience of language comprehension and social cognition. Evidence from neuroimaging and neuropsychological studies has indicated that the representation of social concepts relies on a brain network composed of the bilateral anterior temporal lobe, temporoparietal junction, dorsomedial prefrontal cortex, posterior cingulate gyrus and adjacent precuneus. Recent neuroimaging studies have indicated that the same brain areas may also support social semantic integration at different levels. Future studies should explore the semantic dimensions of social concept representation, the specific components of social semantic processing, and their impacts on human behaviors.

Keywords: concept representation, semantic integration, language comprehension, social cognition